

# A characterization of fine words over a finite alphabet

Amy Glen

*LaCIM, Université du Québec à Montréal, C.P. 8888, succursale Centre-ville,  
Montréal, Québec, CANADA, H3C 3P8*

---

## Abstract

To any infinite word  $\mathbf{t}$  over a finite alphabet  $\mathcal{A}$  we can associate two infinite words  $\min(\mathbf{t})$  and  $\max(\mathbf{t})$  such that any prefix of  $\min(\mathbf{t})$  (resp.  $\max(\mathbf{t})$ ) is the *lexicographically* smallest (resp. greatest) amongst the factors of  $\mathbf{t}$  of the same length. We say that an infinite word  $\mathbf{t}$  over  $\mathcal{A}$  is *fine* if there exists an infinite word  $\mathbf{s}$  such that, for any lexicographic order,  $\min(\mathbf{t}) = a\mathbf{s}$  where  $a = \min(\mathcal{A})$ . In this paper, we characterize fine words; specifically, we prove that an infinite word  $\mathbf{t}$  is fine if and only if  $\mathbf{t}$  is either a *strict episturmian word* or a strict “skew episturmian word”. This characterization generalizes a recent result of G. Pirillo, who proved that a fine word over a 2-letter alphabet is either an (aperiodic) *Sturmian word*, or an ultimately periodic (but not periodic) infinite word, all of whose factors are (finite) Sturmian.

*Key words:* combinatorics on words; lexicographic order; episturmian word; Sturmian word; Arnoux-Rauzy sequence; skew word.

MSC (2000): 68R15.

---

## 1 Introduction

To any infinite word  $\mathbf{t}$  over a finite alphabet  $\mathcal{A}$  we can associate two infinite words  $\min(\mathbf{t})$  and  $\max(\mathbf{t})$  such that any prefix of  $\min(\mathbf{t})$  (resp.  $\max(\mathbf{t})$ ) is the *lexicographically* smallest (resp. greatest) amongst the factors of  $\mathbf{t}$  of the same length (see Pirillo [14]). In the recent paper [15], Pirillo defined *fine words* over two letters; specifically, an infinite word  $\mathbf{t}$  over a 2-letter alphabet  $\{a, b\}$  ( $a < b$ ) is said to be *fine* if  $(\min(\mathbf{t}), \max(\mathbf{t})) = (a\mathbf{s}, b\mathbf{s})$  for some infinite word  $\mathbf{s}$ . Pirillo [15] characterized these words, and remarked that perhaps his characterization can be generalized to an arbitrary finite alphabet; we do just that in this paper. Firstly, we extend the definition of a fine word to more

---

*Email address:* amy.glen@gmail.com (Amy Glen).

than two letters. That is, we say that an infinite word  $\mathbf{t}$  over  $\mathcal{A}$  is *fine* if there exists an infinite word  $\mathbf{s}$  such that, for any lexicographic order,  $\min(\mathbf{t}) = as$  where  $a = \min(\mathcal{A})$ . Roughly speaking, our main result states that an infinite word  $\mathbf{t}$  is fine if and only if  $\mathbf{t}$  is either a *strict episturmian word* or a strict “skew episturmian word” (i.e., a particular kind of non-recurrent infinite word, all of whose factors are *finite episturmian*).

## 2 Notation and terminology

### Finite and infinite words

Let  $\mathcal{A}$  denote a finite alphabet. A (finite) *word* over  $\mathcal{A}$  is an element of the *free monoid*  $\mathcal{A}^*$  generated by  $\mathcal{A}$ , in the sense of concatenation. The identity  $\varepsilon$  of  $\mathcal{A}^*$  is called the *empty word*, and the *free semigroup*, denoted by  $\mathcal{A}^+$ , is defined by  $\mathcal{A}^+ := \mathcal{A}^* \setminus \{\varepsilon\}$ .

Given  $w = x_1x_2 \cdots x_m \in \mathcal{A}^+$  with each  $x_i \in \mathcal{A}$ , the *length* of  $w$  is  $|w| = m$  (note that  $|\varepsilon| = 0$ ). The *reversal*  $\tilde{w}$  of  $w$  is given by  $\tilde{w} = x_mx_{m-1} \cdots x_1$ , and if  $w = \tilde{w}$ , then  $w$  is called a *palindrome*.

An *infinite word* (or simply *sequence*)  $\mathbf{x}$  is a sequence indexed by  $\mathbb{N}$  with values in  $\mathcal{A}$ , i.e.,  $\mathbf{x} = x_0x_1x_2 \cdots$  with each  $x_i \in \mathcal{A}$ . The set of all infinite words over  $\mathcal{A}$  is denoted by  $\mathcal{A}^\omega$ , and we define  $\mathcal{A}^\infty := \mathcal{A}^* \cup \mathcal{A}^\omega$ . An *ultimately periodic* infinite word can be written as  $uv^\omega = uvvv \cdots$ , for some  $u, v \in \mathcal{A}^*$ ,  $v \neq \varepsilon$ . If  $u = \varepsilon$ , then such a word is *periodic*. An infinite word that is not ultimately periodic is said to be *aperiodic*.

A finite word  $w$  is a *factor* of  $z \in \mathcal{A}^\infty$  if  $z = uwv$  for some  $u \in \mathcal{A}^*$ ,  $v \in \mathcal{A}^\infty$ . Further,  $w$  is called a *prefix* (resp. *suffix*) of  $z$  if  $u = \varepsilon$  (resp.  $v = \varepsilon$ ), and we write  $w \prec_p z$  (resp.  $w \prec_s z$ ). We say that  $w$  is a *proper factor* (resp. *prefix*, *suffix*) of  $z$  if  $uv \neq \varepsilon$  (resp.  $v \neq \varepsilon$ ,  $u \neq \varepsilon$ ). An infinite word  $\mathbf{x} \in \mathcal{A}^\omega$  is called a *suffix* of  $\mathbf{z} \in \mathcal{A}^\omega$  if there exists a word  $w \in \mathcal{A}^*$  such that  $\mathbf{z} = w\mathbf{x}$ . A factor  $w$  of a word  $z \in \mathcal{A}^\infty$  is *right* (resp. *left*) *special* if  $wa, wb$  (resp.  $aw, bw$ ) are factors of  $z$  for some letters  $a, b \in \mathcal{A}$ ,  $a \neq b$ .

For  $\mathbf{x} \in \mathcal{A}^\omega$ ,  $F(\mathbf{x})$  denotes the set of all its factors, and  $F_n(\mathbf{x})$  denotes the set of all factors of  $\mathbf{x}$  of length  $n \in \mathbb{N}$ , i.e.,  $F_n(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}^n$ . Moreover, the *alphabet* of  $\mathbf{x}$  is  $\text{Alph}(\mathbf{x}) := F(\mathbf{x}) \cap \mathcal{A}$ , and we denote by  $\text{Ult}(\mathbf{x})$  the set of all letters occurring infinitely often in  $\mathbf{x}$ . Any two infinite words  $\mathbf{x}, \mathbf{y} \in \mathcal{A}^\omega$  are said to be *equivalent* if  $F(\mathbf{x}) = F(\mathbf{y})$ , i.e., if  $\mathbf{x}$  and  $\mathbf{y}$  have the same set of factors. A factor of an infinite word  $\mathbf{x}$  is *recurrent* in  $\mathbf{x}$  if it occurs infinitely many times in  $\mathbf{x}$ , and  $\mathbf{x}$  itself is said to be *recurrent* if all of its factors are recurrent in it.

## Lexicographic order

Suppose the alphabet  $\mathcal{A}$  is totally ordered by the relation  $<$ . Then we can totally order  $\mathcal{A}^*$  by the *lexicographic order*  $<$ , defined as follows. Given two words  $u, v \in \mathcal{A}^+$ , we have  $u < v$  if and only if either  $u$  is a proper prefix of  $v$  or  $u = xau'$  and  $v = xbv'$ , for some  $x, u', v' \in \mathcal{A}^*$  and letters  $a, b$  with  $a < b$ . This is the usual alphabetic ordering in a dictionary, and we say that  $u$  is *lexicographically less* than  $v$ . This notion naturally extends to  $\mathcal{A}^\omega$ , as follows. Let  $\mathbf{u} = u_0u_1u_2\cdots$  and  $\mathbf{v} = v_0v_1v_2\cdots$ , where  $u_j, v_j \in \mathcal{A}$ . We define  $\mathbf{u} < \mathbf{v}$  if there exists an index  $i \geq 0$  such that  $u_j = v_j$  for all  $j = 0, \dots, i-1$  and  $u_i < v_i$ . Naturally,  $\leq$  will mean  $<$  or  $=$ .

Let  $w \in \mathcal{A}^\omega$  and let  $k$  be a positive integer. We denote by  $\min(w|k)$  (resp.  $\max(w|k)$ ) the lexicographically smallest (resp. greatest) factor of  $w$  of length  $k$  for the given order (where  $|w| \geq k$  for  $w$  finite). If  $w$  is infinite, then it is clear that  $\min(w|k)$  and  $\max(w|k)$  are prefixes of the respective words  $\min(w|k+1)$  and  $\max(w|k+1)$ . So we can define, by taking limits, the following two infinite words (see [14])

$$\min(w) = \lim_{k \rightarrow \infty} \min(w|k) \quad \text{and} \quad \max(w) = \lim_{k \rightarrow \infty} \max(w|k).$$

## Morphisms and the free group

A *morphism on  $\mathcal{A}$*  is a map  $\psi : \mathcal{A}^* \rightarrow \mathcal{A}^*$  such that  $\psi(uv) = \psi(u)\psi(v)$  for all  $u, v \in \mathcal{A}^*$ . It is uniquely determined by its image on the alphabet  $\mathcal{A}$ . All morphisms considered in this paper will be non-erasing: the image of any non-empty word is never empty. Hence the action of a morphism  $\psi$  on  $\mathcal{A}^*$  naturally extends to infinite words; that is, if  $\mathbf{x} = x_0x_1x_2\cdots \in \mathcal{A}^\omega$ , then  $\psi(\mathbf{x}) = \psi(x_0)\psi(x_1)\psi(x_2)\cdots$ .

The free monoid  $\mathcal{A}^*$  can be naturally embedded within a free group. We denote by  $\mathcal{F}(\mathcal{A})$  the free group over  $\mathcal{A}$  that properly contains  $\mathcal{A}$ , and is obtained from  $\mathcal{A}$  by adjoining the inverse  $a^{-1}$  of each letter  $a \in \mathcal{A}$ . More precisely, we construct a new alphabet  $\mathcal{A}^\pm$  that consists of all letters  $a$  of  $\mathcal{A}$  and their ‘inverses’  $a^{-1}$ , i.e.,  $\mathcal{A}^\pm = \{a, a^{-1} \mid a \in \mathcal{A}\}$ . If one defines on the free monoid  $(\mathcal{A}^\pm)^*$  the involution  $(a^{-1})^{-1} = a$  for each  $a \in (\mathcal{A}^\pm)^*$ , then necessarily, we have  $(uv)^{-1} = v^{-1}u^{-1}$  for all  $u, v \in (\mathcal{A}^\pm)^*$ . The free group  $\mathcal{F}(\mathcal{A})$  over  $\mathcal{A}$  is the quotient of  $(\mathcal{A}^\pm)^*$  under the relation:  $aa^{-1} = a^{-1}a = \varepsilon$  for all  $a \in \mathcal{A}$ . In what follows, we use the notation  $p^{-1}w$  and  $ws^{-1}$  to indicate the removal of a prefix  $p$  (resp. suffix  $s$ ) from a finite word  $w$ .

Any morphism  $\psi$  on  $\mathcal{A}$  can be uniquely extended to an endomorphism of  $\mathcal{F}(\mathcal{A})$  by defining  $\psi(a^{-1}) = (\psi(a))^{-1}$  for each  $a \in \mathcal{A}$ .

### 3 Episturmian words

An interesting generalization of *Sturmian words* (i.e., aperiodic infinite words of minimal complexity) to a finite alphabet is the family of *Arnoux-Rauzy sequences*, the study of which began in [2] (also see [10,17] for example). More recently, a slightly wider class of infinite words, aptly called *episturmian words*, was introduced by Droubay, Justin, and Pirillo [4] (also see [5,9,11,12] for instance). An infinite word  $\mathbf{t} \in \mathcal{A}^\omega$  is *episturmian* if  $F(\mathbf{t})$  is closed under reversal and  $\mathbf{t}$  has at most one right (or equivalently left) special factor of each length. Moreover, an episturmian word is *standard* if all of its left special factors are prefixes of it. Sturmian words are exactly the aperiodic episturmian words over a 2-letter alphabet.

Standard episturmian words were characterized in [4] using the concept of the *palindromic right-closure*  $w^{(+)}$  of a finite word  $w$ , which is the (unique) shortest palindrome having  $w$  as a prefix (see [3]). Specifically, an infinite word  $\mathbf{s} \in \mathcal{A}^\omega$  is standard episturmian if and only if there exists an infinite word  $\Delta(\mathbf{s}) = x_1x_2x_3 \dots (x_i \in \mathcal{A})$ , called the *directive word* of  $\mathbf{s}$ , such that the infinite sequence of palindromic prefixes  $u_1 = \varepsilon, u_2, u_3, \dots$  of  $\mathbf{s}$  (which exists by results in [4]) is given by

$$u_{n+1} = (u_n x_n)^{(+)}, \quad n \in \mathbb{N}^+. \quad (1)$$

This characterization extends to the case of an arbitrary finite alphabet a construction given in [3] for all *standard Sturmian words*. An important point is that a standard episturmian word  $\mathbf{s}$  can be constructed as a limit of an infinite sequence of its palindromic prefixes, i.e.,  $\mathbf{s} = \lim_{n \rightarrow \infty} u_n$ .

*Note.* Episturmian words are (uniformly) recurrent [4].

#### 3.1 Relation with episturmian morphisms

Let  $a \in \mathcal{A}$  and denote by  $\Psi_a$  the morphism on  $\mathcal{A}$  defined by

$$\Psi_a : \begin{cases} a \mapsto a \\ x \mapsto ax \quad \text{for all } x \in \mathcal{A} \setminus \{a\}. \end{cases}$$

Together with the permutations of the alphabet, all of the morphisms  $\Psi_a$  generate by composition the monoid of *epistandard morphisms* (‘epistandard’ is an elegant shortcut for ‘standard episturmian’ due to Richomme [16]). The submonoid generated by the  $\Psi_a$  only is the monoid of *pure epistandard morphisms*, which includes the *identity morphism*  $\text{Id}_{\mathcal{A}} = \text{Id}$ , and consists of all the *pure standard (Sturmian) morphisms* when  $|\mathcal{A}| = 2$ .

When viewed as an endomorphism of the free group  $\mathcal{F}(\mathcal{A})$ , the morphism  $\Psi_a$  is invertible; that is,  $\Psi_a$  is a *positive automorphism* of  $\mathcal{F}(\mathcal{A})$ , and its inverse is given by

$$\Psi_a^{-1} : \begin{cases} a \mapsto a \\ x \mapsto a^{-1}x \quad \text{for all } x \in \mathcal{A} \setminus \{a\}. \end{cases}$$

It follows that every epistandard morphism is a (positive) automorphism of  $\mathcal{F}(\mathcal{A})$ . See [6,8,16,18] for work involving the invertibility of episturmian morphisms.

**Remark 3.1** If  $\mathbf{x} = \Psi_a(\mathbf{y})$  or  $\mathbf{x} = a^{-1}\Psi_a(\mathbf{y})$  for some  $\mathbf{y} \in \mathcal{A}^\omega$  and  $a \in \mathcal{A}$ , then the letter  $a$  is *separating* for  $\mathbf{x}$  and its factors; that is, any factor of  $\mathbf{x}$  of length 2 contains the letter  $a$ .

Another useful characterization of standard episturmian words is the following (see [9]). An infinite word  $\mathbf{s} \in \mathcal{A}^\omega$  is standard episturmian with directive word  $\Delta(\mathbf{s}) = x_1x_2x_3 \cdots$  ( $x_i \in \mathcal{A}$ ) if and only if there exists an infinite sequence of recurrent infinite words  $\mathbf{s}^{(0)} = \mathbf{s}$ ,  $\mathbf{s}^{(1)}$ ,  $\mathbf{s}^{(2)}$ ,  $\dots$  such that  $\mathbf{s}^{(i-1)} = \Psi_{x_i}(\mathbf{s}^{(i)})$  for all  $i \in \mathbb{N}^+$ . Moreover, each  $\mathbf{s}^{(i)}$  is a standard episturmian word with directive word  $\Delta(\mathbf{s}^{(i)}) = x_{i+1}x_{i+2}x_{i+3} \cdots$ , the *i-th shift* of  $\Delta(\mathbf{s})$ .

To the prefixes of the directive word  $\Delta(\mathbf{s}) = x_1x_2 \cdots$ , we associate the morphisms

$$\mu_0 := \text{Id}, \quad \mu_n := \Psi_{x_1}\Psi_{x_2} \cdots \Psi_{x_n}, \quad n \in \mathbb{N}^+,$$

and define the words

$$h_n := \mu_n(x_{n+1}), \quad n \in \mathbb{N},$$

which are clearly prefixes of  $\mathbf{s}$ . For the palindromic prefixes  $(u_i)_{i \geq 1}$  given by (1), we have the following useful formula [9]

$$u_{n+1} = h_{n-1}u_n;$$

whence, for  $n > 1$  and  $0 < p < n$ ,

$$u_n = h_{n-2}h_{n-3} \cdots h_1h_0 = h_{n-2}h_{n-3} \cdots h_{p-1}u_p. \quad (2)$$

**Remark 3.2** Evidently, if a standard episturmian word  $\mathbf{s}$  begins with the letter  $x \in \mathcal{A}$ , then  $x$  is separating for  $\mathbf{s}$  (see [4, Lemma 4]).

### 3.2 Strict episturmian words

A standard episturmian word  $\mathbf{s} \in \mathcal{A}^\omega$ , or any equivalent (episturmian) word, is said to be  *$\mathcal{B}$ -strict* (or *k-strict* if  $|\mathcal{B}| = k$ , or *strict* if  $\mathcal{B}$  is understood) if  $\text{Alph}(\Delta(\mathbf{s})) = \text{Ult}(\Delta(\mathbf{s})) = \mathcal{B} \subseteq \mathcal{A}$ . In particular, a standard episturmian word over  $\mathcal{A}$  is  *$\mathcal{A}$ -strict* if every letter in  $\mathcal{A}$  occurs infinitely often in its directive

word. The  $k$ -strict episturmian words have complexity  $(k - 1)n + 1$  for each  $n \in \mathbb{N}$ ; such words are exactly the  $k$ -letter Arnoux-Rauzy sequences. Note that the 2-strict episturmian words correspond to the (aperiodic) Sturmian words.

**Remark 3.3** Suppose  $\mathbf{s} \in \mathcal{A}^\omega$  is a standard episturmian word. If  $\mathbf{s}$  is not  $\mathcal{A}$ -strict, then  $\text{Ult}(\Delta(\mathbf{s})) = \mathcal{B} \subset \mathcal{A}$  and there exists a  $\mathcal{B}$ -strict standard episturmian word  $\mathbf{s}'$  and a pure epistandard morphism  $\mu$  on  $\mathcal{A}$  such that  $\mathbf{s} = \mu(\mathbf{s}')$ . More precisely, let  $\Delta(\mathbf{s}) = x_1x_2x_3\cdots$  and let  $m$  be minimal such that  $\text{Alph}(x_{m+1}x_{m+2}\cdots) = \mathcal{B} \subset \mathcal{A}$ . That is,  $x_1x_2\cdots x_m$  is the shortest prefix of  $\Delta(\mathbf{s})$  that contains all the letters not appearing infinitely often in  $\Delta(\mathbf{s})$ , namely the letters in  $\mathcal{A} \setminus \mathcal{B}$ . Then  $\mathbf{s} = \mu_m(\mathbf{s}^{(m)})$  where  $\mathbf{s}^{(m)}$  is the  $\mathcal{B}$ -strict standard episturmian word with directive word  $\Delta(\mathbf{s}^{(m)}) = x_{m+1}x_{m+2}\cdots$ . For example, if  $\Delta(\mathbf{s}) = c(ab)^\omega$ , then  $\mathbf{s} = \Psi_c(\mathbf{s}^{(1)})$  where  $\Delta(\mathbf{s}^{(1)}) = (ab)^\omega$ , i.e.,  $\mathbf{s}^{(1)}$  is the well-known *Fibonacci word* over  $\{a, b\}$ .

## 4 Fine words

Recall that an infinite word  $\mathbf{t}$  over  $\mathcal{A}$  is *fine* if there exists an infinite word  $\mathbf{s}$  such that, for any lexicographic order,  $\min(\mathbf{t}) = a\mathbf{s}$  where  $a = \min(\mathcal{A})$ .

*Note.* Since there are only two lexicographic orders on words over a 2-letter alphabet, a fine word  $\mathbf{t}$  over  $\{a, b\}$  ( $a < b$ ) satisfies  $(\min(\mathbf{t}), \max(\mathbf{t})) = (a\mathbf{s}, b\mathbf{s})$  for some infinite word  $\mathbf{s}$ .

Recently, Pirillo [15] characterized fine words over a 2-letter alphabet. Specifically:

**Proposition 4.1** [15] *Suppose  $\mathbf{t}$  is an infinite word over  $\{a, b\}$ . Then the following properties are equivalent:*

- i)  $\mathbf{t}$  is fine;
- ii) either  $\mathbf{t}$  is a Sturmian word, or  $\mathbf{t} = v\mu(x)^\omega$  where  $\mu$  is a pure standard Sturmian morphism on  $\{a, b\}$ , and  $v$  is a non-empty suffix of  $\mu(x^py)$  for some  $p \in \mathbb{N}$  and  $x, y \in \{a, b\}$  ( $x \neq y$ ).

In other words, a fine word over two letters is either a Sturmian word or an ultimately periodic (but not periodic) infinite word, all of whose factors are (finite) Sturmian, i.e., a so-called *skew Sturmian* word (see [13]). In this paper, we generalize Pirillo's result to infinite words over two or more letters.

The next two propositions are needed for the proof of our main result (Theorem 4.6, to follow). Recall that the Arnoux-Rauzy sequences are precisely the strict episturmian words.

**Proposition 4.2** [10] *Suppose  $\mathbf{s}$  is an infinite word over a finite alphabet  $\mathcal{A}$ . Then the following properties are equivalent:*

- i)  $\mathbf{s}$  is a standard Arnoux-Rauzy sequence;
- ii)  $a\mathbf{s} = \min(\mathbf{s})$  for any letter  $a \in \mathcal{A}$  and lexicographic order  $<$  satisfying  $a = \min(\mathcal{A})$ .

**Proposition 4.3** [14] *Suppose  $\mathbf{s}$  is an infinite word over a finite alphabet  $\mathcal{A}$ . Then the following properties are equivalent:*

- i)  $\mathbf{s}$  is standard episturmian;
- ii)  $a\mathbf{s} \leq \min(\mathbf{s})$  for any letter  $a \in \mathcal{A}$  and lexicographic order  $<$  satisfying  $a = \min(\mathcal{A})$ .

The following key lemma is also needed. From now on, it will be convenient to denote by  $\mathbf{v}_p$  the prefix of length  $p$  of a given infinite word  $\mathbf{v}$ .

**Lemma 4.4** *Let  $\mathcal{A}$  be a finite alphabet and let  $a \in \mathcal{A}$ . Suppose  $\mathbf{t}, \mathbf{s} \in \mathcal{A}^\omega$  are infinite words such that  $\mathbf{t} = \Psi_z(\mathbf{t}^{(1)})$  and  $\mathbf{s} = \Psi_z(\mathbf{s}^{(1)})$  for some  $z \in \text{Alph}(\mathbf{t}^{(1)})$ . Then*

$$\min(\mathbf{t}^{(1)}) = a\mathbf{s}^{(1)} \quad \Leftrightarrow \quad \min(\mathbf{t}) = a\mathbf{s}.$$

**Remark 4.5** Let  $\mathbf{t}, \mathbf{t}^{(1)}, \mathbf{s}, \mathbf{s}^{(1)} \in \mathcal{A}^\omega$  be such that  $\mathbf{t} = \Psi_z(\mathbf{t}^{(1)})$  and  $\mathbf{s} = \Psi_z(\mathbf{s}^{(1)})$  for some letter  $z$  (not necessarily in  $\text{Alph}(\mathbf{t}^{(1)})$ ). Using similar reasoning as in the proof below, it can be shown that

$$\min(\mathbf{t}^{(1)}) = a\mathbf{s}^{(1)} \quad \Leftrightarrow \quad \min(\mathbf{t}) = \begin{cases} zas & \text{if } z < a, \\ as & \text{if } z \geq a. \end{cases}$$

For example, let  $\mathcal{A} = \{a, b, c\}$  with  $a < b < c$  and suppose  $\mathbf{f}$  is the Fibonacci word over  $\{a, b\}$  (i.e., the standard episturmian word directed by  $(ab)^\omega$ ). Then  $\min(\mathbf{f}) = a\mathbf{f}$ , and hence  $\min(\Psi_c(\mathbf{f})) = a\Psi_c(\mathbf{f})$ . On the other hand, if  $\mathbf{f}'$  is the Fibonacci word over  $\{b, c\}$ , then  $\min(\mathbf{f}') = b\mathbf{f}'$  and we have  $\min(\Psi_a(\mathbf{f}')) = ab\Psi_a(\mathbf{f}')$ . Lemma 4.4 is a special case of this result with  $z \in \text{Alph}(\mathbf{t}^{(1)}) \subseteq \mathcal{A}$  and is sufficient for our purposes.

#### Proof of Lemma 4.4

( $\Leftarrow$ ): We have  $\min(\mathbf{t}) = a\mathbf{s}$ . First observe that  $a \in \text{Alph}(\mathbf{t}^{(1)})$ . Indeed, if  $a = z$ , then  $a \in \text{Alph}(\mathbf{t}^{(1)})$  since  $z \in \text{Alph}(\mathbf{t}^{(1)})$  (in fact,  $zz \in F(\mathbf{t})$  since  $zz$  is a prefix of  $a\mathbf{s} = z\mathbf{s}$ , and hence  $a = z \in \text{Alph}(\mathbf{t}^{(1)})$ ). On the other hand, if  $a \neq z$ , then we must have  $a \in \text{Alph}(\mathbf{t}^{(1)})$ , otherwise  $a$  is not in the alphabet of  $\mathbf{t} = \Psi_z(\mathbf{t}^{(1)})$ , which is impossible since  $F(a\mathbf{s}) \subseteq F(\mathbf{t})$ .

Now we show that  $F(a\mathbf{s}^{(1)}) \subseteq F(\mathbf{t}^{(1)})$ . Suppose not, i.e., suppose  $F(a\mathbf{s}^{(1)}) \not\subseteq F(\mathbf{t}^{(1)})$ .

$F(\mathbf{t}^{(1)})$ . Then there exists a minimal  $m \in \mathbb{N}^+$  such that  $as_m^{(1)} \notin F(\mathbf{t}^{(1)})$ . Therefore, if  $\mathbf{s}_m^{(1)} = \mathbf{s}_{m-1}^{(1)}x$  where  $x \in \mathcal{A}$ , then

$$as_{m-1}^{(1)}x \notin F(\mathbf{t}^{(1)}).$$

Letting  $\mathbf{s}_l = \Psi_z(\mathbf{s}_{m-1}^{(1)})$ , we have

$$a\Psi_z(\mathbf{s}_{m-1}^{(1)}x) = a\mathbf{s}_l\Psi_z(x) \in F(as) \subseteq F(\mathbf{t}),$$

and hence  $\Psi_z(a)\mathbf{s}_l\Psi_z(x) \in F(\mathbf{t})$  since  $z$  is separating for  $\mathbf{t}$ . So, if  $x \neq z$  then  $as_{m-1}^{(1)}x \in F(\mathbf{t}^{(1)})$ , which is impossible; whence  $x = z$ . But then,  $\mathbf{s}_{m+1}^{(1)} = \mathbf{s}_{m-1}^{(1)}zy'$  for some  $y' \in \mathcal{A}$  and we have

$$a\Psi_z(\mathbf{s}_{m+1}^{(1)}) = a\mathbf{s}_l z \Psi_z(y') \in F(as) \subseteq F(\mathbf{t}).$$

Thus,

$$\Psi_z(a)\mathbf{s}_l z z \in F(\mathbf{t}),$$

and hence  $as_{m-1}^{(1)}z (= as_{m-1}^{(1)}x)$  is a factor of  $\mathbf{t}^{(1)}$ ; a contradiction. Therefore, we conclude that  $F(as^{(1)}) \subseteq F(\mathbf{t}^{(1)})$ .

Now suppose on the contrary that  $\min(\mathbf{t}^{(1)}) \neq as^{(1)}$ . Then there exists a word  $w^{(1)} \in F(\mathbf{t}^{(1)})$  of minimal length  $|w^{(1)}| = m$  such that

$$w^{(1)} < as_{m-1}^{(1)}.$$

Let  $w^{(1)} = u^{(1)}x$  ( $x \in \mathcal{A}$ ) where  $u^{(1)}$  is non-empty since  $a \in \text{Alph}(t^{(1)})$ . Then, by minimality of  $m$ ,  $u^{(1)} \geq as_{m-2}^{(1)}$ , and therefore  $u^{(1)} = as_{m-2}^{(1)}$  (otherwise  $u^{(1)} > as_{m-2}^{(1)}$  implies  $w^{(1)} > as_{m-1}^{(1)}$ ). Hence,

$$w^{(1)} = as_{m-2}^{(1)}x \quad \text{with } w^{(1)} < as_{m-1}^{(1)},$$

and therefore

$$\mathbf{s}_{m-1}^{(1)} = \mathbf{s}_{m-2}^{(1)}y \quad \text{for some } y \in \mathcal{A}, y > x.$$

Now, letting  $w = \Psi_z(w^{(1)})$  and  $\mathbf{s}_l = \Psi_z(\mathbf{s}_{m-2}^{(1)})$ , we have

$$w = \Psi_z(a)\mathbf{s}_l\Psi_z(x) \in F(\mathbf{t}) \quad \text{and} \quad \mathbf{s}_l\Psi_z(y) \prec_p \mathbf{s}.$$

Now consider  $x', y' \in \mathcal{A}$  such that  $w^{(1)}x' \in F(\mathbf{t}^{(1)})$  and  $\mathbf{s}_m^{(1)} = \mathbf{s}_{m-2}^{(1)}yy'$  is a prefix of  $\mathbf{s}^{(1)}$ . Then  $\Psi_z(w^{(1)}x')$  is a factor of  $\mathbf{t}$ , where

$$\Psi_z(w^{(1)}x') = w\Psi_z(x') = \Psi_z(a)\mathbf{s}_l\Psi_z(x)\Psi_z(x') = \begin{cases} \Psi_z(a)\mathbf{s}_l z \Psi_z(x') & \text{if } x = z, \\ \Psi_z(a)\mathbf{s}_l z x \Psi_z(x') & \text{if } x \neq z. \end{cases}$$



Therefore, the word  $v = as_lzx$  is a factor of  $\mathbf{t}$ . Moreover,  $\Psi_z(\mathbf{s}_m^{(1)})$  is a prefix of  $\mathbf{s}$ , where

$$\Psi_z(\mathbf{s}_m^{(1)}) = \begin{cases} \mathbf{s}_l z \Psi_z(y') & \text{if } y = z, \\ \mathbf{s}_l z y \Psi_z(y') & \text{if } y \neq z, \end{cases}$$

and hence  $\mathbf{s}_{l+2} = \mathbf{s}_l z y$  is a prefix of  $\mathbf{s}$ . Accordingly,  $v < as_{l+2}$  since  $x < y$ , contradicting the fact that the prefixes of  $as$  are the lexicographically smallest factors of  $\mathbf{t}$ . Thus, we conclude that  $\min(\mathbf{t}^{(1)}) = as^{(1)}$ .

( $\Rightarrow$ ): We have  $\min(\mathbf{t}^{(1)}) = as^{(1)}$ . As above, it is easily shown that  $F(as) \subseteq F(\mathbf{t})$ ; whence  $\min(\mathbf{t}) \leq as$ . Let us suppose  $\min(\mathbf{t}) \neq as$ . Then there exists a word  $w \in F(\mathbf{t})$  of minimal length  $|w| = l$  such that  $w < as_{l-1}$ . If we let  $w = ux$ ,  $x \in \mathcal{A}$ , then  $u \geq as_{l-2}$ , and hence  $u = as_{l-2}$  (otherwise  $w > as_{l-1}$ ). Therefore,

$$w = as_{l-2}x < as_{l-1}$$

and hence  $\mathbf{s}_{l-1} = \mathbf{s}_{l-2}y$  where  $y \in \mathcal{A}$ ,  $y > x$ .

Since the letter  $z$  is separating for  $\mathbf{t}$ ,  $\mathbf{s}_{l-2}$  must end with  $z$ ; otherwise  $x = y = z$ , which is impossible. Thus,

$$w = as_{l-3}zx \quad \text{and} \quad \mathbf{s}_{l-1} = \mathbf{s}_{l-3}zy, \quad y > x.$$

Let  $\mathbf{s}_{m-1}^{(1)} = \Psi_z^{-1}(\mathbf{s}_{l-1})$ . We distinguish two cases:  $y = z$  and  $y \neq z$ .

*Case 1:  $y = z$ .* We have  $\mathbf{s}_{l-1} = \mathbf{s}_{l-3}zz$ , and thus  $\mathbf{s}_{m-1}^{(1)}$  and  $\mathbf{s}_{m-2}^{(1)}$  both end with the letter  $z$ . Note that  $z \neq a$  because  $a \leq x < y = z$ . Therefore, since  $w$  begins with  $a$  and  $z$  is separating for  $\mathbf{t}$ , we have  $zw \in F(\mathbf{t})$ . Now,

$$\begin{aligned} \Psi_z^{-1}(zw) &= \Psi_z^{-1}(zas_{l-1}z^{-1}x) \\ &= as_{m-1}^{(1)}z^{-1}\Psi_z^{-1}(x) \\ &= as_{m-1}^{(1)}z^{-1}z^{-1}x \\ &= as_{m-3}^{(1)}x, \end{aligned}$$

i.e.,  $zw = \Psi_z(w^{(1)})$  where  $w^{(1)} = as_{m-3}^{(1)}x \in F(\mathbf{t}^{(1)})$ . Therefore, as  $\mathbf{s}_{m-2}^{(1)}$  ends with  $z > x$ , we have  $w^{(1)} < as_{m-2}^{(1)}$ ; a contradiction.

*Case 2:  $y \neq z$ .* In this case,  $\mathbf{s}_l = \mathbf{s}_{l-3}zyz = \mathbf{s}_{l-1}z$ , and so  $\mathbf{s}_m^{(1)} = \mathbf{s}_{m-1}^{(1)}y' = \mathbf{s}_{m-2}^{(1)}yy'$  for some  $y' \in \mathcal{A}$ . If  $z \neq a$ , then  $zw = zas_{l-3}zx$  is a factor of  $\mathbf{t}$  since  $w \in F(\mathbf{t})$  and  $z$  is separating for  $\mathbf{t}$ . So, letting  $w' = zw$  if  $z \neq a$  and  $w' = w$  if

$z = a$ , we have  $w' \in F(\mathbf{t})$  and

$$\begin{aligned}\Psi_z^{-1}(w') &= a\Psi_z^{-1}(\mathbf{s}_{l-3}zx) \\ &= a\Psi_z^{-1}(\mathbf{s}_{l-1}y^{-1}x) \\ &= a\mathbf{s}_{m-1}^{(1)}(z^{-1}y)^{-1}\Psi_z^{-1}(x) \\ &= a\mathbf{s}_{m-1}^{(1)}y^{-1}z\Psi_z^{-1}(x) \\ &= a\mathbf{s}_{m-2}^{(1)}z\Psi_z^{-1}(x).\end{aligned}$$

That is,  $w' = \Psi_z(w^{(1)})$  where  $w^{(1)} \in F(\mathbf{t}^{(1)})$  is given by

$$w^{(1)} = \begin{cases} a\mathbf{s}_{m-2}^{(1)}xx & \text{if } x = z, \\ a\mathbf{s}_{m-2}^{(1)}x & \text{if } x \neq z. \end{cases}$$

Therefore, since  $\mathbf{s}_m^{(1)} = \mathbf{s}_{m-1}^{(1)}y' = \mathbf{s}_{m-2}^{(1)}yy'$  with  $y > x$ , we have  $w^{(1)} < a\mathbf{s}_{m-1}^{(1)}$ ; a contradiction.

Both Cases 1 and 2 lead to a contradiction; whence  $\min(\mathbf{t}) = a\mathbf{s}$ .  $\square$

We now prove our main result: a characterization of fine words over a finite alphabet. (Recall that  $\mathbf{v}_p$  denotes the prefix of length  $p$  of a given infinite word  $\mathbf{v}$ , and  $\tilde{\mathbf{v}}_p$  denotes its reversal.)

**Theorem 4.6** *Suppose  $\mathbf{t}$  is an infinite word with  $\text{Alph}(\mathbf{t}) = \mathcal{A}$ . Then,  $\mathbf{t}$  is fine if and only if one of the following holds:*

- i)  $\mathbf{t}$  is a strict episturmian word;
- ii)  $\mathbf{t} = v\mu(\mathbf{v})$  where  $\mathbf{v}$  is a  $\mathcal{B}$ -strict standard episturmian word with  $\mathcal{B} = \mathcal{A} \setminus \{x\}$ ,  $\mu$  is a pure epistandard morphism on  $\mathcal{A}$ , and  $v$  is a non-empty suffix of  $\mu(\tilde{\mathbf{v}}_p x)$  for some  $p \in \mathbb{N}$ .

**PROOF.** In what follows, let  $\mathcal{A}$  denote the alphabet of  $\mathbf{t}$ .

( $\Rightarrow$ ):  $\mathbf{t}$  is fine, so there exists an infinite word  $\mathbf{s}$  such that, for any letter  $a \in \mathcal{A}$  and lexicographic order  $<$  satisfying  $a = \min(\mathcal{A})$ , we have  $\min(\mathbf{t}) = a\mathbf{s}$ . Further,  $\min(\mathbf{t}) \leq \min(\mathbf{s})$  since  $F(\mathbf{s}) \subseteq F(a\mathbf{s}) \subseteq F(\mathbf{t})$ , and therefore  $a\mathbf{s} \leq \min(\mathbf{s})$ . Thus, Proposition 4.3 implies that  $\mathbf{s}$  is a standard episturmian word over  $\mathcal{A}$ . We distinguish two cases, below.

*Case 1:*  $a\mathbf{s} = \min(\mathbf{s})$  for any  $a \in \mathcal{A}$  and lexicographic order such that  $a = \min(\mathcal{A})$ .

By Proposition 4.2,  $\mathbf{s}$  is an  $\mathcal{A}$ -strict standard episturmian word.

Clearly,  $F(\mathbf{s}) \subseteq F(\mathbf{t})$  and we show that  $F(\mathbf{s}) = F(\mathbf{t})$  which implies  $\mathbf{t}$  is equivalent to  $\mathbf{s}$ , and hence  $\mathbf{t}$  is an  $\mathcal{A}$ -strict episturmian word. Suppose, on the

contrary,  $F(\mathbf{s}) \neq F(\mathbf{t})$ . Then there exists a word  $u \in F(\mathbf{t}) \setminus F(\mathbf{s})$ , say  $u = xv$  with  $|v| = m$  minimal and  $x \in \mathcal{A}$ . Now,  $v$  is not a prefix of  $\mathbf{s}$ ; otherwise,  $zv$  is a factor of  $\mathbf{s}$  for all  $z \in \mathcal{A}$  (since any prefix of  $\mathbf{s}$  is left special and has  $|\mathcal{A}|$  distinct left extensions in  $\mathbf{s}$ ), which contradicts the fact that  $xv \notin F(\mathbf{s})$ . Therefore, for some order such that  $x = \min(\mathcal{A})$ , we have  $u = xv < x\mathbf{s}_m$ , contradicting the fact that  $\min(\mathbf{t}) = x\mathbf{s}$ ; whence  $F(\mathbf{t}) = F(\mathbf{s})$ .

*Case 2:*  $x\mathbf{s} < \min(\mathbf{s})$  for some  $x \in \mathcal{A}$  and lexicographic order such that  $x = \min(\mathcal{A})$ .

In this case, it follows from Propositions 4.2 and 4.3 that  $\mathbf{s} \in \mathcal{A}^\omega$  is a standard episturmian word that is not  $\mathcal{A}$ -strict. Therefore, letting  $\Delta(\mathbf{s}) = x_1x_2x_3\cdots$ , there exists a minimal  $n \in \mathbb{N}$  such that  $\mathbf{s} = \Psi_{x_1}\Psi_{x_2}\cdots\Psi_{x_n}(\mathbf{s}^{(n)}) = \mu_n(\mathbf{s}^{(n)})$ , where  $\mathbf{s}^{(n)}$  is a  $\mathcal{B}$ -strict standard episturmian word with  $\mathcal{B} = \text{Alph}(\Delta(\mathbf{s}^{(n)})) = \text{Alph}(x_{n+1}x_{n+2}x_{n+3}\cdots) \subset \mathcal{A}$  (see Remark 3.3). Note that if  $\mathbf{s} = \mathbf{s}^{(0)}$  is  $\mathcal{B}$ -strict with  $\mathcal{B} \subset \mathcal{A}$ , then  $n = 0$  and we take  $x_n = x_0$  to be a letter in  $\mathcal{A} \setminus \mathcal{B}$ .

Clearly,  $\mathbf{s}$  begins with  $x_1 \in \mathcal{A}$  and  $x_1$  is separating for  $\mathbf{s}$ . Observe that  $x_1$  must also be separating for  $\mathbf{t}$ . Indeed, let us suppose that this is not true. Then, there exist letters  $z, z' \in \mathcal{A} \setminus \{x_1\}$  (possibly equal) such that  $zz' \in F(\mathbf{t})$ . But, if  $<$  is an order such that  $\min(\mathcal{A}) = z \leq z' < x_1$ , then  $zx_1$  is a prefix of  $z\mathbf{s}$  with  $zz' < zx_1$ , contradicting the fact that  $\min(\mathbf{t}) = z\mathbf{s}$ . Therefore,  $x_1$  must be separating for  $\mathbf{t}$ .

Now, let  $<$  be an order with  $a = \min(\mathcal{A})$ . Let  $\mathbf{t}' = \mathbf{t}$  if  $\mathbf{t}$  begins with  $x_1$ . Otherwise, if  $\mathbf{t}$  begins with  $y \neq x_1$ , let  $\mathbf{t}' = x_1\mathbf{t}$ . In the latter case,  $ax_1 \prec_p a\mathbf{s}$  and  $x_1y \prec_p \mathbf{t}'$  with  $ax_1 < x_1y$ ; thus  $\min(\mathbf{t}') = \min(\mathbf{t}) = a\mathbf{s}$ . So we may consider  $\mathbf{t}'$  instead of  $\mathbf{t}$ .

Observe that  $\mathbf{s} = \Psi_{x_1}(\mathbf{s}^{(1)})$  and, since  $x_1$  is separating for  $\mathbf{s}$  (and hence for  $\mathbf{t}'$ ), we have  $\mathbf{t}' = \Psi_{x_1}(\mathbf{t}^{(1)})$  for some  $\mathbf{t}^{(1)} \in \mathcal{A}^\omega$ . Because  $\min(\mathbf{t}') = x\mathbf{s}$  for any letter  $x \in \mathcal{A}$  and lexicographic order such that  $x = \min(\mathcal{A})$ , it follows that  $\text{Alph}(\mathbf{t}^{(1)}) = \mathcal{A}$  (see arguments in the first lines of the proof of Lemma 4.4); in particular  $x_1 \in \text{Alph}(\mathbf{t}^{(1)})$ . So, by Lemma 4.4, we have

$$\min(\mathbf{t}^{(1)}) = a\mathbf{s}^{(1)}.$$

Continuing in the same way (and applying Lemma 4.4 repeatedly), we obtain sequences  $(\mathbf{s}^{(i)})$ ,  $(\mathbf{t}^{(i)})$ ,  $(\mathbf{t}'^{(i)})$  for  $i = 0, 1, 2, \dots, n$  such that  $\mathbf{s}^{(i-1)} = \Psi_{x_i}(\mathbf{s}^{(i)})$ ,  $\mathbf{t}'^{(i-1)} = \Psi_{x_i}(\mathbf{t}^{(i)})$ , where  $\mathbf{t}'^{(i-1)} = \mathbf{t}^{(i-1)}$  if  $\mathbf{t}^{(i-1)}$  begins with  $x_i$ ,  $\mathbf{t}'^{(i-1)} = x_i\mathbf{t}^{(i-1)}$  otherwise, and  $\mathbf{t}^{(0)} = \mathbf{t}$ ,  $\mathbf{t}'^{(0)} = \mathbf{t}'$ . In particular, we have  $\text{Alph}(\mathbf{t}^{(n)}) = \mathcal{A}$  and

$$\min(\mathbf{t}^{(n)}) = a\mathbf{s}^{(n)}$$

for any  $a \in \mathcal{A}$  and lexicographic order  $<$  satisfying  $a = \min(\mathcal{A})$ .

Now we show that  $\mathcal{B} = \mathcal{A} \setminus \{x_n\}$ , i.e.,  $\mathcal{A} = \mathcal{B} \cup \{x_n\}$ . First observe that  $x_n \in \mathcal{A} \setminus \mathcal{B}$  by minimality of  $n$ .

Suppose  $\mathbf{t}^{(n)}$  contains two occurrences of the letter  $x_n$ . Then, since  $x_{n+1}$  is separating for  $\mathbf{t}^{(n)}$ , we have  $x_n w^{(n)} x_n \in F(\mathbf{t}^{(n)})$  for some non-empty word  $w^{(n)}$  for which  $x_{n+1}$  is separating, and the first and last letter of  $w^{(n)}$  is  $x_{n+1}$  (that is,  $w^{(n)} x_n = \Psi_{x_{n+1}}(w^{(n+1)} x_n)$ , where  $w^{(n+1)} = \Psi_{x_{n+1}}^{-1}(w^{(n)} x_{n+1}^{-1})$ ). Continuing the above procedure, we obtain infinite words  $\mathbf{t}^{(n+1)}, \mathbf{t}^{(n+2)}, \dots$  containing similar shorter factors  $x_n w^{(n+1)} x_n, x_n w^{(n+2)} x_n, \dots$  until we reach  $\mathbf{t}^{(q)}$ , which contains  $x_n x_n$ . But this is impossible because  $x_{q+1} \in \mathcal{B} \subseteq \mathcal{A} \setminus \{x_n\}$  is separating for  $\mathbf{t}^{(q)}$ . Therefore,  $\mathbf{t}^{(n)}$  contains only one occurrence of  $x_n$  and we have

$$\mathbf{t}^{(n)} = u x_n \mathbf{v} \quad \text{for some } u \in (\mathcal{A} \setminus \{x_n\})^* \text{ and } \mathbf{v} \in (\mathcal{A} \setminus \{x_n\})^\omega.$$

Note that the same reasoning allows to prove the unicity of  $x_0 \in \mathcal{A} \setminus \mathcal{B}$  when  $n = 0$ .

Clearly, for any order such that  $x_n = \min(\mathcal{A})$ , we have  $\min(\mathbf{t}^{(n)}) = x_n \mathbf{v} = x_n \mathbf{s}^{(n)}$ ; whence  $\mathbf{v} = \mathbf{s}^{(n)}$  and so  $\mathbf{t}^{(n)} = u x_n \mathbf{s}^{(n)}$ . Note that if  $u \neq \varepsilon$ , then  $u$  ends with  $x_{n+1}$ , and in particular  $x_{n+1}$  is separating for  $u x_n$  since  $x_{n+1}$  is separating for  $\mathbf{t}^{(n)}$ .

Let  $u' = x_{n+1} u$  if  $u$  does not begin with  $x_{n+1}$ ; otherwise let  $u' = u$ . Then  $u' x_n$  is a prefix of  $\mathbf{t}^{(n)}$ . Moreover, since  $x_{n+1}$  is separating for  $u' x_n$ , we have  $u' x_n = \Psi_{x_{n+1}}(u^{(n+1)} x_n)$  where  $u^{(n+1)} = \Psi_{x_{n+1}}^{-1}(u' x_{n+1}^{-1})$ . Hence  $\mathbf{t}^{(n+1)} = u^{(n+1)} x_n \mathbf{s}^{(n+1)}$ , where  $x_{n+2}$  is separating for  $u^{(n+1)} x_n$  (if  $u^{(n+1)} \neq \varepsilon$ ). Continuing in this way, we arrive at the infinite word  $\mathbf{t}^{(q)} = x_n \mathbf{s}^{(q)}$  for some  $q \geq n$ .

Now, reversing the procedure, we find that

$$\mathbf{t}^{(n)} = w \mathbf{s}^{(n)} \quad \text{where } w = u x_n \text{ is a non-empty suffix of } \Psi_{x_{n+1}} \cdots \Psi_{x_q}(x_n).$$

Accordingly,  $u \in \mathcal{B}^*$  since  $x_{n+1}, \dots, x_q \in \mathcal{B}$ ; whence  $\mathcal{A} = \mathcal{B} \cup \{x_n\}$ .

Suppose  $(u_i)_{i \geq 1}$  is the sequence of palindromic prefixes of  $\mathbf{s}$  and the words  $(h_i)_{i \geq 0}$  are the prefixes  $(\mu_i(x_{i+1}))_{i \geq 0}$  of  $\mathbf{s}$ . Then, letting  $u_i^{(n)}, h_i^{(n)}$ , and  $\mu_i^{(n)}$  denote the analogous elements for  $\mathbf{s}^{(n)}$ , we have

$$\mu_0^{(n)} = \text{Id}, \quad \mu_i^{(n)} = \Psi_{x_{n+1}} \Psi_{x_{n+2}} \cdots \Psi_{x_{n+i}} = \mu_n^{-1} \mu_{n+i}$$

and

$$h_0^{(n)} = x_{n+1}, \quad h_i^{(n)} = \mu_i^{(n)}(x_{n+1+i}) \quad \text{for } i = 1, 2, \dots$$

Now, if  $u \neq \varepsilon$ , then  $q \geq n + 1$ , and we have

$$\begin{aligned}
\Psi_{x_{n+1}} \cdots \Psi_{x_q}(x_n) &= \mu_{q-n}^{(n)}(x_n) = \mu_{q-n-1}^{(n)} \Psi_q(x_n) \\
&= \mu_{q-n-1}^{(n)}(x_q x_n) \\
&= h_{q-n-1}^{(n)} \mu_{q-n-1}^{(n)}(x_n) \\
&\quad \vdots \\
&= h_{q-n-1}^{(n)} \cdots h_1^{(n)} \mu_0^{(n)}(x_{n+1} x_n) \\
&= h_{q-n-1}^{(n)} \cdots h_1^{(n)} h_0^{(n)} x_n = u_{q-n+1}^{(n)} x_n \quad (\text{by (2)}).
\end{aligned}$$

Therefore,  $w = ux_n$  where  $u$  is a (possibly empty) suffix of the palindromic prefix  $u_{q-n+1}^{(n)}$  of  $\mathbf{s}^{(n)}$ . That is,  $u$  is the reversal of some prefix of  $\mathbf{s}^{(n)} = \mathbf{v}$ ; in particular

$$u = \tilde{\mathbf{v}}_p \quad \text{for some } p \in \mathbb{N},$$

and hence

$$\mathbf{t}^{(n)} = \tilde{\mathbf{v}}_p x_n \mathbf{v}.$$

So, passing back from  $\mathbf{t}^{(n)}$  to  $\mathbf{t}$ , we find that

$$\mathbf{t} = v\mu_n(\mathbf{v}) = v\mathbf{s} \quad \text{where } v \text{ is a non-empty suffix of } \mu_n(\tilde{\mathbf{v}}_p x_n).$$

Cases 1 and 2 give properties *i*) and *ii*), respectively.

( $\Leftarrow$ ): Firstly, if  $\mathbf{t}$  is an  $\mathcal{A}$ -strict episturmian word, then Proposition 4.2 implies that  $\mathbf{t}$  is fine.

Now suppose  $\mathbf{t} = v\mu(\mathbf{v})$  where  $\mathbf{v}$  is a  $\mathcal{B}$ -strict standard episturmian word with  $\mathcal{B} = \mathcal{A} \setminus \{x\}$ ,  $\mu$  is a pure epistandard morphism on  $\mathcal{A}$ , and  $v$  is a non-empty suffix of  $\mu(\tilde{\mathbf{v}}_p x)$  for some  $p \in \mathbb{N}$ . First observe that if  $\mu = \text{Id}$ , then

$$\mathbf{t} = \tilde{\mathbf{v}}_q x \mathbf{v} \quad \text{for some } q \leq p.$$

Consider an order  $<$  such that  $\min(\mathcal{A}) = a \neq x$ . Then, by Proposition 4.2,  $\min(\mathbf{v}) = a\mathbf{v}$ , and it follows that  $\min(\mathbf{t}) = \min(\mathbf{v}) = a\mathbf{v}$ . Indeed, if  $\mathbf{t} = x\mathbf{v}$  (i.e.,  $q = 0$ ), then it is clear that  $\min(\mathbf{t}) = \min(\mathbf{v})$ . On the other hand, if  $q \geq 1$ , let us suppose, on the contrary, that  $\min(\mathbf{t}) \neq \min(\mathbf{v})$ . Then  $\min(\mathbf{t})$  is a suffix of  $\tilde{\mathbf{v}}_q x \mathbf{v}$  containing the letter  $x$ , i.e.,

$$\min(\mathbf{t}) = a\tilde{\mathbf{v}}_l x \mathbf{v} \quad \text{for some } l \text{ with } 0 \leq l < q \leq p$$

(where  $a\tilde{\mathbf{v}}_l = \tilde{\mathbf{v}}_{l+1} \in F(\mathbf{v})$ ). But, since  $\min(\mathbf{v}) = a\mathbf{v}$ ,  $a\mathbf{v}_{l+1}$  is a factor of  $\mathbf{v}$  (and hence a factor of  $\mathbf{t}$ ) with  $a\mathbf{v}_{l+1} = a\mathbf{v}_l a < a\tilde{\mathbf{v}}_l x$ ; a contradiction. Thus  $\min(\mathbf{t}) = \min(\mathbf{v}) = a\mathbf{v}$ . Moreover, it is clear that  $\min(\mathbf{t}) = x\mathbf{v}$  for any order

such that  $x = \min(\mathcal{A})$ . So we have shown that  $\min(\mathbf{t}) = a\mathbf{v}$  for any letter  $a \in \mathcal{A}$  and lexicographic order satisfying  $a = \min(\mathcal{A})$ ; whence  $\mathbf{t}$  is fine.

Now consider the case when  $\mu$  is not the identity. Let us suppose that  $\mathbf{t}$  is not fine and let  $\mu$  be minimal with this property. Then,  $\mu = \Psi_z\eta$  for some  $z \in \mathcal{A}$  and pure epistandard morphism  $\eta$ .

Consider  $\mathbf{t}' = v'\mu(\mathbf{v})$ , where  $v' = v$  if  $v$  begins with  $z$  and  $v' = zv$  otherwise. Then  $v'$  is also a non-empty suffix of  $\mu(\tilde{\mathbf{v}}_p x)$  since  $z$  is separating for the word  $\mu(\tilde{\mathbf{v}}_p x)$  (which begins with  $z$ ). Letting  $\mathbf{t}^{(1)} = \Psi_z^{-1}(\mathbf{t}')$ , we have

$$\mathbf{t}^{(1)} = w\eta(\mathbf{v})$$

where  $w = \Psi_z^{-1}(v')$  is a non-empty suffix of  $\eta(\tilde{\mathbf{v}}_p x)$ . By minimality of  $\mu$ ,  $\mathbf{t}^{(1)}$  is fine, so there exists an infinite word  $\mathbf{s}^{(1)} = \Psi_z^{-1}(\mathbf{s})$  such that  $\min(\mathbf{t}^{(1)}) = a\mathbf{s}^{(1)}$  for any  $a \in \mathcal{A}$  and order  $<$  satisfying  $a = \min(\mathcal{A})$ . But then  $\min(\mathbf{t}) = \min(\mathbf{t}') = a\mathbf{s}$  by Lemma 4.4. Thus  $\mathbf{t}$  is fine.  $\square$

**Example 4.7** Let  $\mathcal{A} = \{a, b, c\}$  with  $a < b < c$  and suppose  $\mathbf{f}$  is the Fibonacci word over  $\{a, b\}$ . Then, the following infinite words are fine.

- $\mathbf{f} = abaababaabaaba \dots$
- $c\mathbf{f} = \underline{c}abaababaabaaba \dots$
- $\tilde{\mathbf{f}}_4 c\mathbf{f} = aaba\underline{c}abaababaabaaba \dots$
- $\Psi_a(\mathbf{f}) = aabaaabaabaaabaaba \dots$
- $\Psi_c(c\mathbf{f}) = \underline{c}cacbcacacbcacbcacacbcacacbc \dots$
- $\Psi_c(\tilde{\mathbf{f}}_4 c\mathbf{f}) = cacacbcac\underline{c}cacbcacacbcacacbcacacbcac \dots$

Let us note, for example, that  $\Psi_c(\mathbf{f})$  is **not** fine since it is a *non-strict* standard episturmian word. That is,  $\Psi_c(\mathbf{f})$  is a standard episturmian word with directive word  $c(ab)^\omega$ , so it is not strict, nor does it take the second form given in Theorem 4.6.

## 5 Concluding remarks

It is easy to see that Proposition 4.1 is a special case of Theorem 4.6 because the 2-strict episturmian words are precisely the Sturmian words and the 1-strict standard episturmian words are periodic infinite words of the form  $x^\omega$  where  $x$  is a letter (see [9, Proposition 2.9]).

As alluded to in the introduction, an infinite word taking form *ii*) in Theorem 4.6 is said to be a strict *skew episturmian* word. Skew episturmian words

(now called *episkew words* [1,7]) are explicated in the paper [7], in which we expand on our work here by characterizing via lexicographic order all *episturmian words in a wide sense*, i.e., all infinite words whose factors are (finite) episturmian.

## Acknowledgements

The author would like to thank Jacques Justin for suggesting the problem and giving many helpful comments on preliminary versions of this paper. Thanks also to the referees for their careful reading of the paper and providing thoughtful suggestions.

## References

- [1] J.-P. Allouche, A. Glen. Extremal properties of (epi)sturmian sequences and distribution modulo 1, in preparation.
- [2] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité  $2n+1$ , *Bull. Soc. Math. France* **119** (1991), 199–215.
- [3] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, *Theoret. Comput. Sci.* **183** (1997), 45–82.
- [4] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* **255** (2001), 539–553.
- [5] A. Glen, Powers in a class of  $\mathcal{A}$ -strict standard episturmian words, *Theoret. Comput. Sci.* **380** (2007), 330–354.
- [6] A. Glen, *On Sturmian and episturmian words, and related topics*, PhD thesis, The University of Adelaide, Australia, April 2006.  
(<http://thesis.library.adelaide.edu.au/public/adt-SUA20060426.164255/index.html>)
- [7] A. Glen, J. Justin, G. Pirillo, Characterizations of finite and infinite episturmian words via lexicographic orderings, *European J. Combin.* (in press), doi:10.1016/j.ejc.2007.01.002.
- [8] Eddie Godelle, *Automorphismes de groupes libres, tresses et morphismes épisturmiens*, Preprint (2006), <http://hal.archives-ouvertes.fr/hal-00022410/en/>.
- [9] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.* **276** (2002), 281–313.
- [10] J. Justin, G. Pirillo, On a characteristic property of Arnoux-Rauzy sequences, *Theor. Inform. Appl.* **36** (2002), 385–388.

- [11] J. Justin, G. Pirillo, Episturmian words: shifts, morphisms and numeration systems, *Internat. J. Found. Comput. Sci.* **15** (2004), 329–348.
- [12] J. Justin, L. Vuillon, Return words in Sturmian and episturmian words, *Theor. Inform. Appl.* **34** (2000), 343–356.
- [13] M. Morse, G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories, *Amer. J. Math.* **62** (1940), 1–42.
- [14] G. Pirillo, Inequalities characterizing standard Sturmian and episturmian words, *Theoret. Comput. Sci.* **341** (2005), 276–292.
- [15] G. Pirillo, Morse and Hedlund’s skew Sturmian words revisited, *Ann. Comb.* (accepted).
- [16] G. Richomme, Conjugacy and episturmian morphisms, *Theoret. Comput. Sci.* **302** (2003), 1–34.
- [17] R.N. Risely, L.Q. Zamboni, A generalization of Sturmian sequences: combinatorial structure and transcendence, *Acta Arith.* **95** (2000), 167–184.
- [18] Z.-W. Wen, Y. Zhang, Some remarks on invertible substitutions on three letter alphabet, *Chinese Sci. Bull.* **44** (1999), 1755–1760.